

Anush Sankaran¹, Naveen Panwar¹, Shreya Khare¹, Senthil Mani¹, Akshay Sethi², Rahul Aralikkat¹, Neelamadhav Gantayat¹

¹IBM Research AI, ²IIIT Delhi, India

Contact Email: anussank@in.ibm.com

Motivation

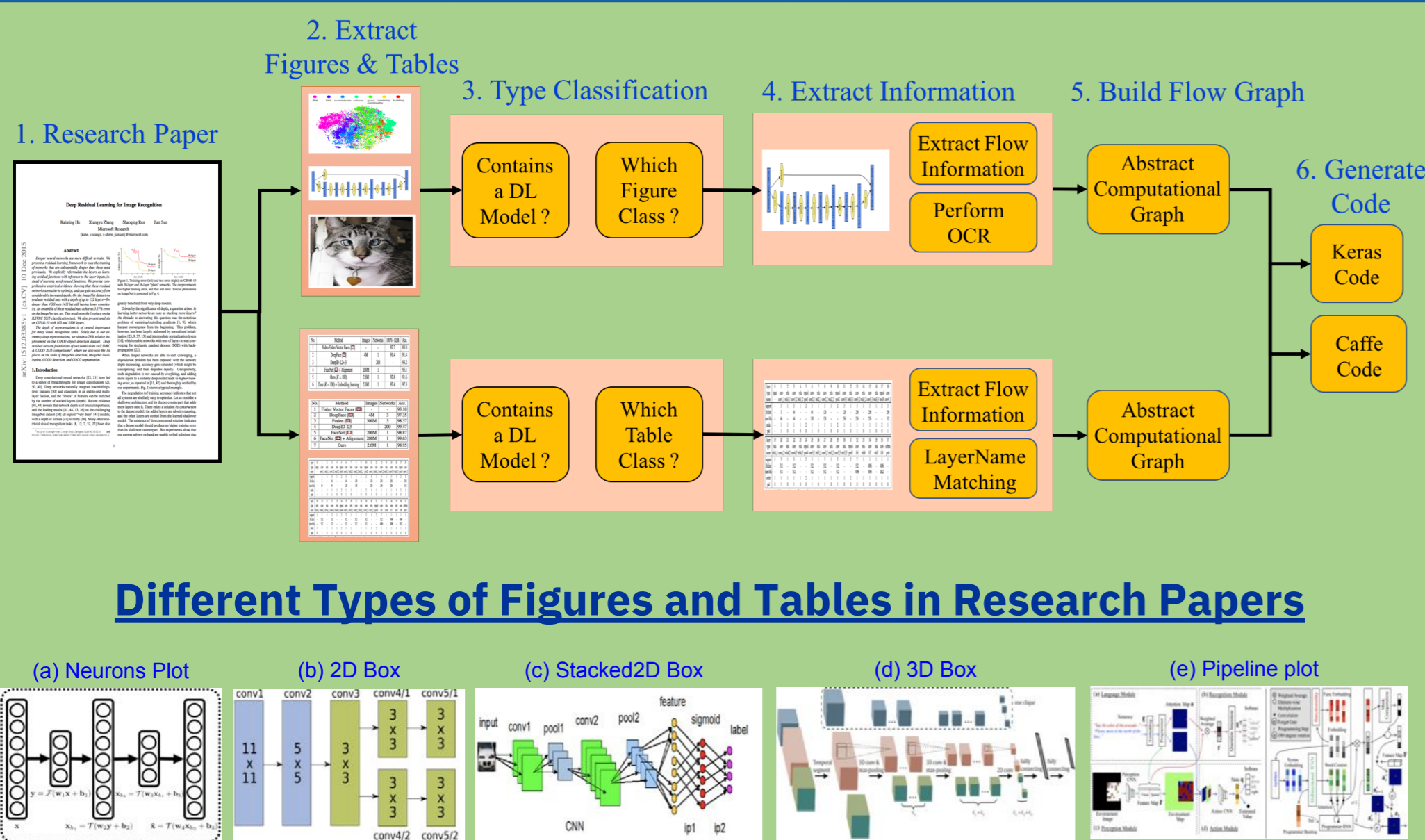
- In NIPS 2016, ~685/2500 were in “Deep Learning or Neural Networks”
- Only **101/567 (~18%)** papers made their source implementation available in the open source domain

Three highly referred papers in “Image Captioning”

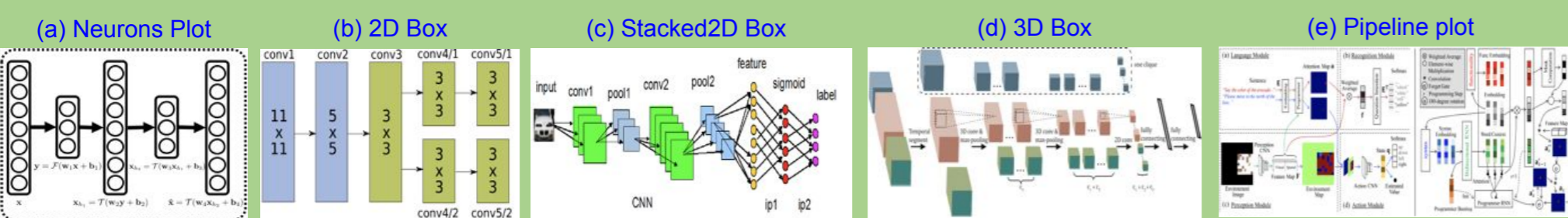
- Show and Tell:** Original implementation available in Theano
- NeuralTalk2:** Original implementation available in Torch
- LRCN:** Implementation available in Caffe

Aim: Reproducibility of existing research works is a challenge. Can we automatically parse papers to generate code?

Proposed Approach

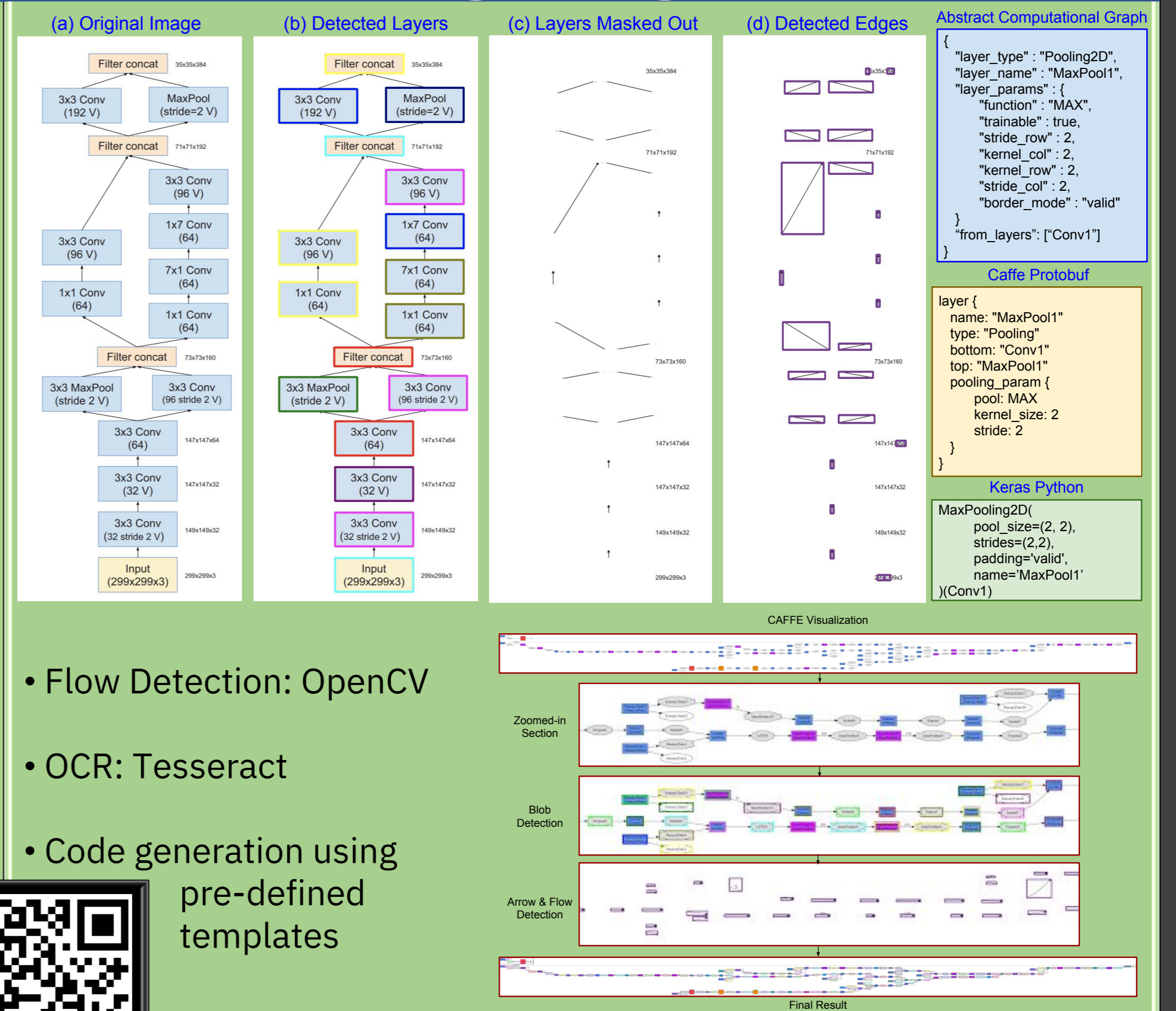


Different Types of Figures and Tables in Research Papers



layer	type	name	strides	kernel	padding	FLIPS
conv1	conv	conv1	220x220x3	110x110x64	7x7x3,2	9K 115M
pool1	pool	pool1	110x110x64	55x55x64	3x3x64,2	0 0
conv2	conv	conv2	55x55x64	55x55x64	1x1x64,1	4K 13M
pool2	pool	pool2	55x55x64	27x27x64	3x3x64,2	111K 355M
conv3	conv	conv3	27x27x64	27x27x128	3x3x128,2	37K 20M
pool3	pool	pool3	27x27x128	13x13x128	3x3x128,2	664K 521M
conv4	conv	conv4	13x13x128	13x13x256	3x3x256,1	148K 173M
conv5	conv	conv5	13x13x256	13x13x256	1x1x256,1	66K 13M
conv6	conv	conv6	13x13x256	13x13x256	3x3x256,1	596K 116M
conv7	conv	conv7	13x13x256	13x13x256	3x3x256,1	0 0
conv8	conv	conv8	13x13x256	13x13x256	3x3x256,1	100M 103M
conv9	conv	conv9	13x13x256	13x13x256	3x3x256,1	34M 34M
conv10	conv	conv10	13x13x256	13x13x256	3x3x256,1	52K 0.5M
conv11	conv	conv11	13x13x256	13x13x256	3x3x256,1	0 0
conv12	conv	conv12	13x13x256	13x13x256	3x3x256,1	140M 1.6M

Extracting Image Content



- Flow Detection: OpenCV
- OCR: Tesseract
- Code generation using pre-defined templates

Simulated Dataset

- Manually define a grammar for generating trainable deep learning models along with hyper-parameters
- 216,000 models** generated with Keras and Caffe visualizations, Keras and Caffe code
- 5 – 40 layers deep models with 3,000 layers per depth

Current Layer	Dense	Conv2D	Flatten	Dropout	Maxpool	AvgPool	Concat	Embed	RNN	RNN (seq)	LSTM	LSTM (seq)
Dense	✓						✓	✓				
Conv2d		✓	✓	✓	✓	✓	✓					
Flatten	✓			✓			✓	✓				
Dropout					Same as previous layer							
MaxPool		✓	✓	✓	✓	✓	✓					
AvgPool		✓	✓	✓	✓	✓	✓					
Concat	If input is one dimensional, same as Dense layer; else same as previous layer											
Embed			✓	✓			✓	✓	✓	✓	✓	✓
RNN	✓			✓	✓		✓	✓	✓	✓	✓	✓
RNN (seq)				✓	✓		✓	✓	✓	✓	✓	✓
LSTM	✓			✓	✓		✓	✓	✓	✓	✓	✓
LSTM (seq)				✓	✓		✓	✓	✓	✓	✓	✓

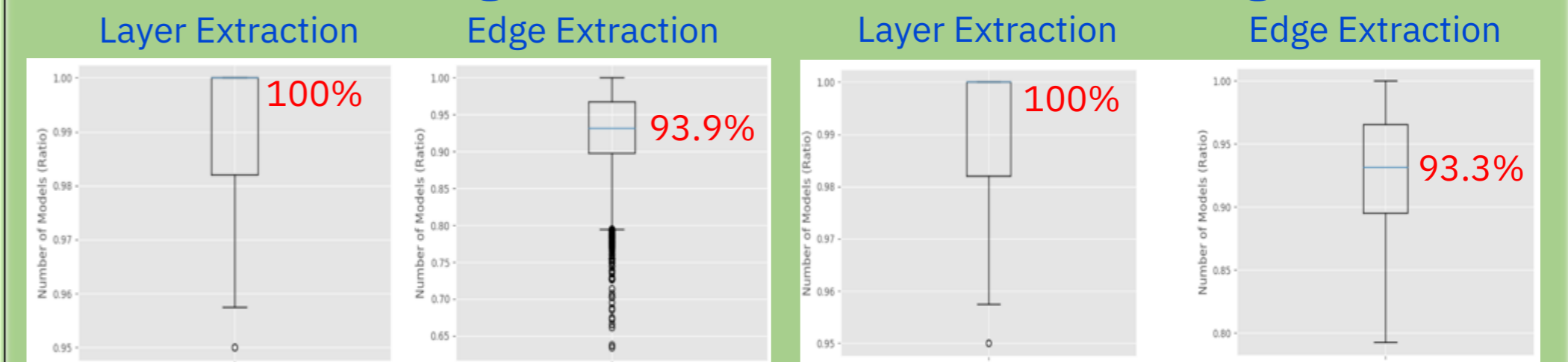
Experimental Results

DL Flow Image classification using VGG19 (fc2) features

Observation	Binary Classifier			Five-class Classifier		
	Train	Val	Test	Train	Val	Test
#DataPoints	18,592	6,197	6,198	1,720	573	574
Naive Bayes	77.29%	64.39%	62.56%	40.42%	54.30%	52.84%
Decision Tree	99.96%	76.67%	74.35%	99.65%	50.57%	49.13%
Logistic Regression	99.96%	86.17%	85.27%	99.65%	69.98%	68.47%
RDF	99.96%	83.72%	82.94%	99.65%	68.72%	66.02%
SVM (RBF Kernel)	99.96%	86.89%	85.25%	99.65%	72.94%	69.68%
Neural Network	99.96%	87.93%	86.25%	100%	74.93%	71.60%

Keras Images

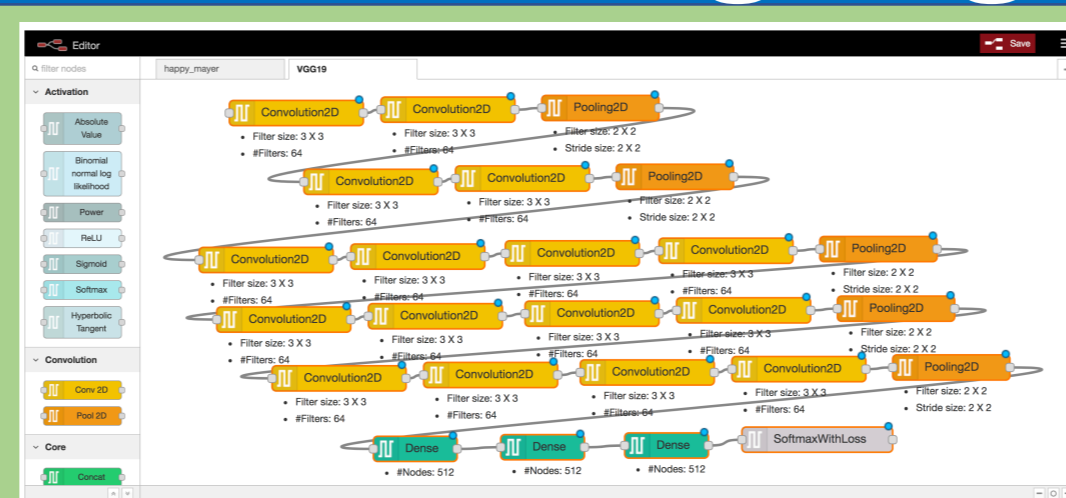
Caffe Images



71.5% accurate in choosing the type of image and 93% accurate in extracting flow information

Discussion & Ongoing Work

- Downloaded 5000 papers from arXiv
- Automatically extract design flow by proposed approach
- Enable crowdsourced contribution using a intuitive drag-and-drop framework



- Extend for other image model types
- Extract hyper-parameter information
- Better fusion of table and image content
- Extract information from descriptive text